DOCUMENT RESUME

ED 142 706 08 CE 011 429

TITLE	Checklist for Reporting Results of Student Outcome Studies in Career Education.
INSTITUTION	New Educational Directions, Crawfordsville, Ind.
SPONS AGENCY	Office of Career Education (DHEW/OE), Washington, D.C.
PUB DATE	May 77
GRANT	G007604329
NOTE	26p.
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	*Achievement Gains; *Career Education; *Evaluation Criteria; Evaluation Methods; *Program Effectiveness; Program Evaluation; Research Criteria; Student Evaluation; Student Improvement; *Technical Writing; *Test Results

ABSTRACT

Intended for use by career education program staff and by evaluators in reviewing draft evaluation reports, this checklist is a guide for ensuring that evaluation reports contain all the information needed by others in interpreting reported results. It applies only to the sections of evaluation reports concerning programs' impact on students and was designed for the typical student outcome evaluation where pencil-and-paper test scores of one or more groups of students are analyzed using inferential statistics. Following presentation of the checklist itself is an elaboration of the rationale and requirements of each checklist item. A fictitious student impact study report is included to demonstrate how the information is associated with each checklist item. (TA)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

NED NEW EDUCATIONAL DIRECTIONS, INC.
a nonprofit service organization
BOX 307 CRAWFORDSVILLE, INDIANA 47933 (317) 362-8877

ED142706

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CHECKLIST FOR REPORTING RESULTS OF STUDENT OUTCOME STUDIES IN CAREER EDUCATION

May 1977

Prepared under Grant Number G007604329 from:

The U. S. Office of Career Education
Office of Education
Department of Health, Education, and Welfare

Project Title: Synthesizing and Communicating Career Education Evaluation Results

Project Director: Deborah G. Bonnet
Senior Research Associate
New Educational Directions, Inc.

CE 011 429

ACKNOWLEDGEMENT

Mr. Terry Newall of the Office of Career Education lent a great deal of assistance in the preparation of this document. His constructively critical reviews of the several draft versions are very much appreciated.

DISCRIMINATION PROHIBITED

Title VI of the Civil Rights Act of 1964 states: "No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance." Title IX of the Education Amendments of 1972, Public Law 92-318, states: "No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance." Therefore, career education projects supported under Sections 402 and 406 of the Education Amendments of 1974, like every program or activity receiving financial assistance from the U.S. Department of Health, Education, and Welfare, must be operated in compliance with these laws.

The material in this publication was prepared pursuant to a grant from the Office of Education, U.S. Department of Health, Education, and Welfare. However, points of view or opinions expressed do not necessarily represent policies or positions of the Office of Education.

CHECKLIST FOR REPORTING RESULTS OF STUDENT OUTCOME STUDIES IN CAREER EDUCATION

What the Checklist Is and Is Not

The checklist is a guide for ensuring that evaluation reports contain all of the information needed by others in interpreting reported results. We recognized the need for such a tool after reviewing a number of evaluation reports of career education programs funded by the U.S. Office of Education in the past and finding that these reports frequently lack such essential information as the number of students involved in the evaluation or the name of the statistic applied to the results. These errors of omission are almost certainly the result of oversight in the majority of cases, so it seemed that a system for checking the drafts of evaluation reports would be useful.

The checklist is not a dramatic breakthrough in the field of career education evaluation. It does not delve into the design, implementation, or analysis phases of evaluation.[1] It places no judgement on the quality of various evaluation strategies but rather deals with the most popular designs, be they good, bad, or indifferent. Nor is the checklist a complete guide to report-writing; it concerns itself with content while leaving style and format to the choice of the author.

The checklist applies only to the sections of evaluation reports concerning programs' impact on students. It was designed for the typical student outcome evaluation where paper-and-pencil test scores of one or more groups of students are analyzed utilizing inferential statistics. If your program is using a more unique evaluation approach, the checklist still may be useful, but some of the items will not apply.

Some will be of the opinion that the checklist calls for too much technical information. As we see it, an evaluation report should be meaningful to anyone who may have an interest in your career education program including such diverse groups as the local school staff, the Chamber of Commerce, the U.S. Office of Education, and educators and non-educators nationwide. Most of these people are not statisticians, but some are, and those with technical expertise will tend to be skeptical of reported results if key technical information is not included in a report. It is usually possible to present the more technical material unobtrusively in parenthetical comments, footnotes, and tables while preserving the report's readability and utility for diverse audiences. Some items included in the checklist are redundant, such as the number of students involved in the study and the degrees of freedom associated with the statistic, but the inclusion of both pieces of information will enhance the credibility of the evaluation in the eyes of many.

---

[1] If this is your concern, refer to the U.S.O.E publications:

A Practical Guide to Measuring Project Impact on Student Achievement by D. Horst, K. Tallmadge, and C. Wood of RMC Research Corporation, 1975, ERIC number ED 106376.

Evaluation and Educational Decision-Making: A Functional Guide to Evaluating Career Education by M. B. Young and R. G. Schuh of Development Associates, Inc., 1975, ERIC number ED 117185

Still other individuals, particularly advocates of stringent research designs, will find the standards of the checklist too low. There is no mention, for example, of tests of the homogeneity of variance. Such omissions should not be taken to mean that such information is considered irrelevant. Rather, the checklist is geared toward improving the reporting of evaluations as they are commonly conducted for career education programs at this time. Thus, the items of the checklist should be viewed as essential, but minimal; exceeding these standards is commendable.

The checklist was designed for use by career education program staff and by evaluators in reviewing draft evaluation reports. If used in this way, a copy of the checklist should be made for each student impact study, usually defined as one scale of a test administeted at one grade level. In this way, independent checks can be made of the thoroughness of the reporting of the evaluation components associated with each outcome objective. As the report is reviewed, a check mark should be placed next to items found in the report; a numbered checklist item should be checked off only if all applicable lettered items under it appear in the report.

In a less direct sense the checklist also may prove useful in earlier stages of evaluation by serving as a guideline for ensuring that evaluation plans include the collection of all data and the performance of all analyses needed for the report, for judging the quality of past reports prepared by the bidders for an evaluation contract, and for outlining or writing evaluation reports.

The career educator who is not also a statistician may find several unfamiliar terms in checklist items 6 and 7. In many cases it still will be possible to recognize the information if it is contained in a report. For example, if a table has a column labelled "t" and numbers appear in it, it is not necessary to know what a t is to know that it is reported. We do, however, advise caution to the non-statistician in concluding that particular statistical data are missing, since many statistical terms go by several names. Thus, apparent omissions of a technical nature should be discussed with the author or other specialist before firm conclusions are reached concerning the report's status of items 6 and 7.

Following the checklist itself is a further elaboration of the rationale and requirements of each checklist item. After that is a fictitious student impact study report designed to demonstrate how the information associated with each checklist item may appear in an actual report.

CHECKLIST FOR REPORTING RESULTS OF
STUDENT OUTCOME STUDIES IN
CAREER EDUCATION

___1. The career education objective(s) assessed by the study

___2. The career education program whose impact is being assessed

- ___a. Conceptual approach
- ___b. Staff training in career education
- ___c. Student activities

___3. The students involved in the study

- ___a. Method of selecting students for each group
- ___b. Grade level
- ___c. Number of students in each group (N)
- ___d. Unique characteristics of students involved in the study
- ___e. If a comparison group is used, evidence that the comparison group is comparable to the career education group

___4. The measurement tool

If you use a commercially-available test, include:

- ___a. The full name of the test
- ___b. The name or number of the form
- ___c. The publisher
- ___d. The name of each scale used in the evaluation
- ___e. A description of the specific skills, attitudes, or knowledge measured by each scale
- ___f. The kind of score analyzed

If you use a locally-developed test, include:

- ___g. The name of the test
- ___h. The name of each scale
- ___i. A description of specific skills, attitudes, or knowledge measured by each scale
- ___j. A copy of the test
- ___k. The scoring key
- ___l. A description of test development procedures
- ___m. Any available information concerning reliability and validity

___5. Test administration

- ___a. Dates of testing
- ___b. Testing procedures
- ___c. Rationale for any elimination of scores before analysis

___6. Descriptive statistical results

___a. Group means

___b. Standard deviations

___7. Inferential statistical parameters and results

If you use a t-test this includes:

___a. Whether the independent groups or matched-pairs t was used

___b. Whether the test was one-tailed or two-tailed

___c. Degrees of freedom (df)

___d. Value of t

___e. Value of p

If you use the analysis of variance or the analysis of covariance this includes:

___f. The analysis of variance design

___g. Degrees of freedom

___h. Value of F

___i. Value of p

___j. If significant differences are found and the study involves three or more groups of students, a test of multiple comparisons

___8. Interpretation of results

___a. The meaning of the statistical results

___b. Your interpretation of the reason for the results

___c. The educational significance of results

___9. Final checks

___a. Clearly labelled tables

___b. Accurately typed numerical data

Explanation of Checklist Items

1. The career education objective(s) assessed by the study

The career education objective provides the rationale for the evaluation. It should be stated in terms of what should be different or better about students as a result of their career education experiences. The same information can be conveyed in the form of evaluation questions or hypotheses.

2. The career education program whose impact is being assessed

Measurement of the extent and nature of career education exposure is for some programs as large a problem as measurement of student outcomes themselves. However, if evaluation results are to have utility either within or outside of the school system, it is critical to describe the career education program as thoroughly and as quantitatively as possible.

a. Conceptual framework

The conceptual framework of the K-12 program is usually described in the body of the project's report rather than in the evaluation report, but it is included as a checklist item because it is important in interpreting outcome results. The description should indicate for each grade level or grade level grouping: 1) the major objective(s) of career education (e.g., the development of self awareness), and 2) the global implementation strategy (e.g., classroom infusion).

b. Staff training in career education

"Staff", as used here, refers to the individuals who "teach" career education to students and may include counselors, librarians, community resource persons, parents, etc. Staff training data have two purposes for student impact studies: 1) as an indication of the resources required to produce the observed student effects, and 2) as evidence that staff are familiar with the career education model which they are presumably implementing.

Some guidelines for reporting staff training data in conjunction with student impact studies are:

1. Data should be presented for the specific individuals who delivered career education to the students included in the study. This means that data such as the number of teachers in the district who have participated in inservice sessions are relevant only if students are selected randomly from all classrooms in the district, or if all students in the district are included. Where students are selected for outcome measurement because they are in classrooms where career education is used extensively, the training of their teachers and other staff who "taught" them career education is of interest; district-wide training data have little to do with those particular students' outcomes. This is not to say that data concerning the extent and nature of the project's staff training program should not be collected and reported where student impact assessment focuses on the "most exposed" students, but that these data will not fulfill the purposes for reporting staff training in conjunction with student outcome studies.

2. The time frame as well as the amount of staff training should be specified (e.g., 25 hours of training during the past two years). If the amount is not presented in standard time units (such as hours or days) it should be translatable to time (such as the number of half-day sessions).

3. The source(s) of data should be indicated.

4. Training strategies and topics also should be addressed, but not necessarily in the evaluation report.

c. Student activities

Like staff training data, information concerning the amount and type of career education experienced by students should be given for the specific students involved in the impact assessment. Again, there are often good reasons for determining the average amount and type of career education the district's students have experienced. However, such data are relevant to measured student outcomes only if students are selected randomly from the school or district for outcome measurement rather than on the basis of their participation in a particular set of career education activities.

The *ideal* evaluation would answer *all* of the following questions at each grade level relevant to the student impact assessment:

1. How many and what proportion of students have been exposed to career education? (This question need be answered only if the student sampling plan allows the inclusion in the "career education" group some students who in fact have experienced no career education.)

2. What types of career education activities have these students experienced? Activities should be categorized according to their primary purpose (e.g., self-awareness, job-seeking skills) as well as their operational nature (e.g., field trips, role-playing).

3. How many and what proportion of students experienced each type of activity?

4. What was the average amount of student exposure to each activity type? The amount of exposure should be presented in time units or the number of occasions on which the activity was experienced, depending on the nature of the strategy.

5. What was the average amount of student exposure to "career education", or to all activity types combined?

Realistically, all but the first question may be difficult to answer unless the career education program being assessed consists of a small number of separate courses or discrete instructional units. With full recognition of the difficulties associated with this checklist item, we recommend that available data be presented in per-student terms and that consideration be given early in the program to incorporating the collection of these types of data into the evaluation plan. A few other pointers on this topic are:

1. The time frame of student activities data should be stated. If at all possible the time frame of activity data should be the same as

the time frame of the program under evaluation. Some assessments address the impact of several years of career education implementation. For example, a three-year-old high school program may be assessed by testing twelfth graders. In this case, student activity data should incorporate the first year's tenth grade activities, the second year's eleventh grade activities, and the third year's twelfth grade activities. If you use a pre-post evaluation design, activity data should cover the period between testing dates.

2. Even if the program is primarily classroom-based, the school is likely to provide other career education activities such as counseling, assembly programs or school-wide career fairs. Since such activities are easily overlooked if student activity data are collected solely from teachers, other data sources should be considered.

3. The data source(s) should be indicated in terms of who provided the information, how, and when.

4. If your evaluation involves a comparison group, it should not be assumed that those students have not experienced career education. Activity data should be collected and reported for comparison, as well as career education students.

5. Very often it is impossible to describe student activities as thoroughly and as quantitatively as you would like, but whatever information you have should be shared with the report's readers, even if it is based only on informal observations of the program.

3. *The students involved in the study*

a. Method of selecting students for each group

If students are drawn "randomly" the description of the sampling technique should include any constraints on the randomness of the sample. For example, "The sample was drawn to give proportionate representation to the academic, vocational, and general curricula," or "Since students could not be taken out of class, only students whose schedules included a study hall were tested." It should also indicate whether students were selected on an individual or classroom basis.

If students are selected on the basis of their participation in a particular class or program, the question is not, "How were students selected for the evaluation?", but rather, "How were students selected for the program?" This question often is equivalent to, "Why does a student get this teacher rather than another?", which is answered by describing the school's classroom assignment practices, or it may be, "Why does a student go to this school rather than another?", which no one is likely to ask. Another issue may be the manner in which one of several similar programs was chosen for assessment, which is usually based on one program's relative intensity and/or on the convenience of data gathering.

b. Grade level

If students from several grade levels are included in a single group, indicate the number or proportion representing each grade level. If the school system does not use the traditional K-12 grade level designations give the group's age and the bases for assigning students to classroom units, modules, or courses.

c. *Number of students in each group (N)*

d. *Unique characteristics of the students*

If the evaluation is of a special program for an identified sub-population within the system such as gifted, handicapped, or vocational students, this, of course, should be stated. Since others will want to use your evaluation results in deciding whether to replicate your efforts in other settings, it also is helpful to indicate other, less notable student characteristics, such as the socio-economic nature of the community and any ways in which the students involved in the evaluation are atypical of the school population (such as predominantly male or below-average achievers).

e. *If a comparison group is used, evidence that the comparison group is comparable to the career education group*

This is needed even if both groups are tested on a pre- and post-test basis and the analysis of covariance model is used for analysis. In this case, a test for significant differences between groups on the pre-test will suffice.

For the post-test with comparison group design it is essential to establish the groups' equivalence on educationally-relevant variables if results are to be taken seriously. Without such evidence, it is impossible to say whether differences between the groups' outcomes are due to the career education program or to differences in the students themselves. Better than nothing but still inadequate evidence of group equivalence is a statement like, "The two groups were drawn from schools serving communities of very similar socio-economic characteristics." A little better is, "The two groups represent the two third-grade classes in a school where students are assigned randomly to classroom units," or "Like the career education group, the comparison students were drawn from the college preparatory curriculum." However, a program committing resources to assessing student impact should seriously consider devoting further effort to establishing the credibility of the comparison group. If the school system has a testing program, scores from recently-administered aptitude or achievement tests can be used to test group equivalence on these dimensions. Grade-point averages can be used in the same way.

4. *The measurement tool*

Widely-recognized problems of measurement in career education make it essential to convey precisely what student attributes were measured. The following information provides a sufficient operational definition of these attributes to allow the reader to draw his or her own conclusions regarding the meaning of the results.

If you use a commercially available test include:

a. *The full name of the test*

b. *The name or number of the form*

(If there are multiple levels or parallel forms of the same test.)

c. *The publisher*

If the publisher is not widely known, the full address is also helpful.

d. *The name of each scale used in the evaluation*

e. *A description of the specific skills, attitudes, or knowledge measured by each scale*

This should resemble the career education objective associated with the scale but is usually more narrowly defined. For example, an objective may read, "students will acquire career decision-making skills," but the description of the scale used in measuring the attainment of the objective may be, ". . . assesses the student's ability to select from a list of job titles the occupations most appropriate for an individual whose interests and personality traits are given." Publishers' test manuals often contain adequate scale descriptions which can be quoted directly. This checklist item may be omitted for tests of basic academic skills if scale names clearly identify the attributes they measure (e.g., reading comprehension).

f. *The kind of score analyzed*

Examples are raw scores and standard scores.

Additional helpful items are: The number of items on each scale; a description of item types (e.g., multiple choice); a sample item from each scale; a summary of the publisher's reported evidence of reliability and validity.

If you use a locally-developed test, include:

g. *The name of the test*

Locally-developed tests should be named so that others can reference them easily.

h. *The name of each scale*

i. *A description of the specific skills, attitudes, or knowledge measured by each scale*

(See 4e)

j. *A copy of the test*

If the test has more than one scale, indicate which items belong to which scales.

k. *The scoring key*

For objective tests this can be indicated on the test booklet. For subjective tests the scoring criteria and procedures should be given in detail.

l. *A description of test development*

Include who was involved in test development (e.g., teachers) and the source of concepts tapped by items (e.g., grade level objectives developed by the project staff).

m. Any available information concerning reliability and validity

A minimum standard for establishing the validity of locally-developed tests is a review by individuals other than the test developers for judging whether the test appears to measure the student attributes it is intended to measure. Such reviews also can focus on factors which influence reliability such as the adequacy of the test's length, its readability, and the appropriateness of response scales. If a review is undertaken, the evaluation report should describe 1) who performed the review, 2) the criteria used in assessing the test, 3) a brief summary of results, and 4) the way in which results were used in revising the test. Highly desirable but sometimes impractical is field-testing the instrument before it is used as an evaluative tool. If this is done, specify at a minimum 1) the number and grade level of students involved in the field-testing, 2) the analyses applied to the data, 3) a brief summary of the results of the analyses, and 4) the manner in which the field-test results were used in revising the test.

The same test results used in the evaluation itself also can be analyzed in a variety of ways for describing psychometric properties of the test. It is beyond the scope of this checklist to discuss the reporting of such analyses; suffice it to say that whatever analyses are performed should be reported.

5. Test administration

a. Dates of testing

Indicate within a week or two the time(s) of data collection (e.g., "the last week in May").

b. Testing procedures

Note should be made of any significant deviations from the administration procedures stipulated by the publisher of commercial tests, of differences in procedures in administration for different groups of students, or of changes in procedures between pre- and post-testing.

c. Rationale for any elimination of scores before analysis

Indicate why and to what extent some students were not included in data analysis. A common reason is missing pre- or post-test scores.

6. Descriptive statistical results

If the inferential statistic is non-parametric, non-parametric descriptive statistics will be substituted for means and standard deviations (e.g., a frequency table for the chi-squared; medians and ranges for the Mann-Whitney U).

a. Groups means

Present all means relevant to the analysis with their associated N's. If the analysis of covariance is applied this includes adjusted post-test means. It is conventional to round means, standard deviations, and values of t and F to the nearest hundredth; more digits than this are unnecessary and confusing. It is also helpful to compute for the reader differences between the means compared in the analysis; present negative differences as negative numbers.

b. Standard deviations

It is good practice to present the standard deviation associated with each mean as they, too, can be interpreted by some in meaningful ways. This item may be omitted so long as the value of t or the complete analysis of variance summary table is presented but it is preferrable to include standard deviations in any event. Be careful to avoid confusion between the standard deviation and the variance.

7. Inferential statistical parameters and results

The guidelines below apply directly only to the most widely-used parametric statistics, but most inferential statistics have parameters analagous to those of the t and F. In any case the report should be very specific about which statistic was applied, including a reference if the statistic is at all uncommon.

If you use a t-test this includes:

a. Whether the independent groups or matched-pairs t was used

Although the appropriate t should be clear from the design, specifying which was used is reassuring.

b. Whether the test was one-tailed or two-tailed

c. Degrees of freedom (df)

d. Value of t

If the difference between means is negative (e.g., the comparison group scored higher than the career education group) t also should be negative.

e. Value of p

Since different people use different standards of statistical significance, it is best to give the observed p value if results meet your standards. If you report "no significant differences", it is important to indicate the alpha level. If, for example, an alpha of .1 is used for three independent analyses, the reported values may be: $p < .1$; $p < .005$; $p > .1$.

The American Physchological Associate (APA) style for presenting items c, d, and e within the text is, "Results indicate that the career education group scored significantly higher than the comparison group, $t\ (48) = 2.62$, $p < .01$."

If you use the analysis of variance or the analysis of covariance this includes:

f. The analysis of variance design

For example, "the one-way analysis of variance for three independent groups" or, "the one-way analysis of covariance, where pre-test scores served as the covariate and post-test scores as the dependent variable."

g. Degrees of freedom (df)

h. Value of F

i. Value of p

(See item 7e)

Although there is a trend away from presenting analysis of variance summary tables, it is still good practice. An adequate format for a one-way analysis of variance (or covariance) is:

Source of Variance	df	MS	F
Between groups	2	125.95	31.25*
Within groups	72	4.03	
Total	74		

*$p < .05$

If the analysis of variance design is specified in the text as suggested above and if the standard deviations associated with each group mean are presented, the analysis of variance summary table may be omitted. The APA style for presenting these results within the text is, "Results of the one-way analysis of variance show significant differences among the mean scores of the three groups, $F(2, 72) = 31.25$, $p < .05$."

Almost any set of data can be analyzed in more than one way. In deciding which analysis or analyses to perform and to report, consider the needs of the various audiences of your evaluation report. Locally, separate analyses for different schools or classrooms may be on interest. This is fine, but audiences outside of the school system will be interested in the finding that "Ms. Richards' class showed improvement but Mr. Thompson's did not," only if the differences in the two teachers' career education strategies are discussed. It is a fairly common practice to analyze outcomes of students at a given grade level separately by school or classroom without identifying in the report any reasons for differing results. Unless the career education experiences of the various groups are different and identified, a summary analysis also should be presented, where all "career educated students" are combined into a single group.

j. If significant differences are found and the study involves three or more groups of students, a test of multiple comparisons

If, for example, the evaluation design includes two groups of students who have experienced different types or amounts of career education and a comparison group exposed to little or no career education, the F statistic alone does not indicate which pair(s) of groups are different, and a multiple comparison test is needed. The statistic and values of its parameters, as well as the results, should be indicated.

8. Interpretation of results

a. The meaning of statistical results

Since many audiences of your report will be unaware of the principles of statistical inference it is important to explain the meaning of the

analysis. Two examples are, "Although the career education students scored on the average somewhat higher than the comparison group, the t test shows that the difference in scores was probably a chance occurrence. Thus, these test results give us no reason to believe that participation in the program influenced the self-esteem of ninth-graders." "The results of the analysis of covariance show that when we account for chance differences between the groups' measured reading achievement prior to the program, we can say with 95% confidence that the superior performance of career education students on the post-test is due to something other than chance. That is, upon completion of the program, career education students read better than we would expect had they not participated in the program."

b. Your interpretation of the reason for the results

An ineffective career education program is only one of many possible explanations of disappointing evaluation results and although it should not be eliminated as a possibility, other factors should be explored. Just a few of these are comparison groups which are unlike career education groups on relevant variables, measurement instruments which are inappropriate for the program or insensitive to instruction and poorly-controlled testing situations. Formulating defensible hypotheses concerning which of these or other factors influenced the results requires familiarity with the full context of the program and the evaluation--a familiarity most of the report's readers will not have. Therefore, your speculation is not only acceptable; it is desirable, so long as speculation is labelled as such.

Similarly, career education experiences rarely constitute the only feasible explanation of positive results. If circumstances would not permit the application of a tightly-controlled evaluation design, addressing the shortcomings of the evaluation is more likely to enhance its credibility than to detract from it. Reinforcing evaluation results with other observations or research is in order especially if you have reason to believe that the results are valid but your evaluation model does not allow conclusive cause-and-effect inferences. If, for example, your evaluation of pre- to post-growth without a reference group yields positive results, the question of whether this growth would have occurred without career education still remains. You may be able to cite other research indicating that at this age youngsters generally grow very little or even regress in this area, as is often the case in the affective domain. Informal observations of teachers, which alone may be inadequate evidence of program impact, also are effective "back-ups" to objective evaluation results. Particularly if these observations are in the form of case histories they make the report both readable and convincing.

Also, if you have reason to believe a particular career education activity or strategy was a major influence on positive results, share this hypothesis.

c. The educational significance of results

Statistical significance can be and often is achieved with a difference between means of one raw score point or less. Thus, you should not assume that statistically significant results alone demonstrate that the program is worth maintaining or expanding. Whereas the judgement of the importance of a given level of student impact to the school system's priorities is more appropriately made by school administrators than by evaluators, the evaluation report can discuss the magnitude of the impact in ways which facilitate these judgements. Although grade equivalent scores and percentile ranks should not

be used for analysis, group means on nationally-normed tests may be converted to these units for this purpose. On other tests it may help to convert raw score means into percentage scores. Another thing to keep in mind is that some career education objectives may seem trivial to some members of your report's audience. If you anticipate this problem, it is advisable to explain the relevance of the objective to the overall goals of career education.

Closely associated with the question of educational significance is that of cost effectiveness. A cost analysis may address any one or combination of the following questions: 1) What did it cost to produce the student impact identified by evaluation results? 2) What would it cost to maintain the program? 3) How does the cost of the program's strategies compare to that of other strategies which produce the same effect? 4) Which cost components could be decreased or eliminated without seriously affecting student impact? It is beyond the scope of the checklist to delve into this topic, but we would like to stress that cost effectiveness analysis adds a very desirable dimension to evaluation and is likely to receive more attention in the future.

This checklist item may be omitted if no statistical significance is achieved. If results are statistically significant educational significance should be addressed. Only if the results are judged educationally significant is it appropriate to consider cost effectiveness.

8. Final checks

a. Clearly labelled tables.

Be sure that tables indicate what the numbers within the table represent.

b. Accurately typed numerical data

Careful proof-reading of final typewritten copy is essential, as conflicting data within a report is common but an easily-avoided problem.

Application of the Checklist

Following is an illustrative student impact study report. The report is entirely fictitious; any resemblance of the program, the setting, the instruments, or the results to anything existing or planned is purely coincidental. The student impact report should be considered just one section of a larger evaluation report which in turn is one section of a project's performance report.

Marginal notations indicate the checklist item or items addressed in each section of the report. The report discusses three different studies as we have defined them and checklist items are subscripted for sections which pertain to only one or two of these studies.

The Fourth-Grade Model Program: Its Impact on Students

3b The student impact assessment at the fourth-grade level focused on three evaluation questions associated with the major elementary objectives of the Career Orientation Project (COP):

1. Does career education affect students' reading achievement? 1_1
2. Does career education affect students' mathematical achievement? 1_2
3. Does career education positively affect students' career awareness? 1_3

Career Education and Comparison Groups

Lincoln Elementary School, judged by the COP staff as the pilot school with the most advanced elementary career education program, served as the experimental site. The administration of Washington Elementary School declined the invitation of last spring to become involved in the project but agreed in December to arrange for the school's fourth graders to serve as a comparison group. The two schools serve adjacent attendance 3a zones of a large suburban community of primarily white-collar workers, and 3d both follow the academic curriculum endorsed by the district. Both schools 3e have a self-contained classroom organization where students are assigned randomly to classroom units. However, the enrollment of Lincoln is about twice that of Washington, with four classroom units at grade four as compared to two fourth-grade classes at Washington.

Since one of the four fourth-grade teachers at Lincoln attended only one COP workshop and considers career education a waste of time his students were not included in the evaluation. The other three teachers have been active in career education since attending COP's orientation program 2b last spring, having participated in the thirty-hour summer workshop and in monthly meetings with the COP elementary consultant.

Each teacher made use of the COP-developed curriculum guide, but the 2a emphasis of various activities varied from class to class, as shown in Table A. The numbers of field trips, guest speakers, and audio visual activities were determined from COP's resource center records of the period of September 7 through May 13; all students present on the days of these activities participated in them. The number of students who 2c shadowed a parent at work sometime during the year was provided by teachers, who determined the number on the basis of students' oral reports.

Two teachers maintained a daily record of career education infusion into content area lessons with a checklist instrument formatted as a

calendar. Infusion was defined as an activity which simultaneously addressed an identifiable content area objective and an identifiable career education objective; thus, the infusion data of Table A partially duplicate the data shown for other strategies. For example, a film about the application of fractions to jobs in the construction industry would be considered a math infusion activity as well as an occupational information audio-visual activity.

Teacher A ascribes to the infusion strategy and, according to the COP elementary consultant, also applies it extensively. However, she found the checklist system burdensome and never submitted her monthly reports. Since the evaluators were not contracted until December, the checklist system did not go into effect until January. The COP staff pointed out that these data for classes B and C may be a slight overestimate of the extent of infusion throughout the school year because the teachers seemed to become more active in career education as the year progressed.

2c

The teachers of the comparison group classes were interviewed by a member of the evaluation team during May in order to determine whether their students had received career education instruction. Both teachers were aware of the COP program and had heard of career education, but neither had participated in any formal training in career education. Both classes had experienced three field trips, but clarifying questions indicated that they were traditional "product-oriented" field trips rather than "career education" field trips. One of the teachers is interested in values clarification and estimated that her class had experienced about one valuing activity per week. Aside from these cases, the comparison students' teachers reported no use of the activities listed in Table A or other conscious efforts toward career orientation in their instruction.

TABLE A

Career Education Activities

of Experimental Group

Strategy	Class		
	A	B	C
Class size (in May)	24	19	23
Values/self awareness AV activities	13	6	7
Occupational information AV activities	3	12	27
Field trips	4	4	4
Guest Speakers	18	13	0
Parent shadows (% students)	78%	61%	95%
Infusion (% daily lessons)			
Language arts	?	24%	17%
Social studies	?	33%	40%
Math	?	21%	23%
Other	?	8%	13%

Academic Achievement

Conveniently, the district's testing program involved the administration of the McDonald Achievement Battery Form GA[1] to all fourth graders in mid-September. The reading comprehension and math concepts scales of this battery, with reported test-retest reliabilities of .82 and .87 respectively, were chosen for analysis.

Two-tailed *t*-tests for independent groups were applied to the pre-test scores of the students enrolled in career education and comparison group classes in September. No significant differences were found between the groups on either scale (reading: $t(103) = -.68$, $p > .1$; math: $t(105) = .19$, $p > .1$). Thus, the analysis of covariance model, which accounts for random pre-test differences, was chosen for determining whether the two groups differed on end-of-year achievement.

The reading comprehension and math concepts scales of the McDonald Form GA were readministered to both groups by their teachers in mid-May. If a student's pre- or post-score on a given scale was missing, he or she

[1] McDonald Test Company, Box 307, Pittsville, Kentucky 44444

was eliminated from the analyses of that scale. Sixty-four career education and forty-six comparison students remained in their respective classes throughout the year; thus, the large majority of eligible students were included in the analyses. Presented in Table B are mean scores, standard deviations, and results of the one-way analysis of covariance, where pre-test scores served as the covariate and post-test scores as the dependent variable.

TABLE B

McDonald Achievement Battery Results
Raw Scores

Scale/Group	N	Mean/ (Std. Dev.) Pre	Post	Adjusted	Analysis of Covariance SV	df	MS	F
READING COMP.								
Career Ed	60	22.64 (6.21)	30.35 (7.08)	30.63	Between	1	18.23	1.67
Comparison	41	23.41 (5.52)	29.67 (5.75)	29.52	Within	98	10.91	$p > .1$
					Total	99		
MATH CONCEPTS								
Career Ed	59	26.11 (7.42)	33.36 (7.63)	33.24	Between	1	52.71	4.23
Comparison	43	25.84 (8.32)	29.76 (8.50)	29.58	Within	99	12.46	$p < .05$
					Total	100		

The mean post-test reading score of the career education group was higher than the comparison group's, but even when we take into consideration the career education group's slightly lower pre-test scores, these end-of-year differences are not statistically significant and are probably due to chance. However, the evaluators learned after becoming involved with COP that the district has a state-wide reputation for its excellent elementary reading program. Achievement scores confirm this reputation; the grade equivalent of the combined groups' means are 4.3 for the September testing and 5.5 in May. Thus, it may have been unreasonable to expect the career education program to improve measurably the already-high-quality reading instruction of these youngsters. It also should be noted that although no positive effects of the inclusion of career education are evident in reading scores, nor are any negative effects.

The career education group outperformed the comparison group on the math computations scale to a degree which the evaluators consider educationally, as well as statistically significant. The difference between the groups' adjusted post-test means slightly exceeds one-third of the test's norm group's standard deviation (3.66/10.31 = .35), a common standard for judging educational significance.[2] Put another way, the grade equivalent of the career education group's year-end mean score was 5.3 as compared to the comparison group's 4.9.

8a2
8c2

There are at least three ways in which the career education program may have resulted in improved math achievement:

1. Because teachers learned more interesting ways to teach math by infusing career education into their lessons, they gave more emphasis to this subject area.
2. By recognizing the relevance of math to their present and future lives, students became more motivated to learn these skills.
3. Through increased "hands-on" experiences in applying math concepts to realistic problems, students were more able to internalize mathematical skills.

8b2

Since the infusion of career education into mathematics was fairly intense and the elementary curriculum guide emphasizes manipulative activities and discussions of "real-world" applications of mathematics, both 2 and 3 above were probably operating. Also, COP's Elementary Consultant reports that many elementary teachers have found themselves devoting more time to math instruction since becoming active in career education, although he does not recall specifically whether this comment was made by Lincoln's fourth-grade teachers.

Career Awareness

3e
4g3
8a3

The program's success in developing students' career awareness was assessed by comparing the career education and comparison groups' year-end performance on the locally-developed Career Quiz. Although no pre-test data were collected with respect to this outcome, the career education and comparison groups' equivalence on socio-economic factors and on

[2] A Practical Guide to Measuring Project Impact on Student Achievement by D. Horst, K. Tallmadge, and C. Wood of RMC Research Corporation, 1975, ERIC number ED 106376, p. 69.

September academic achievement make it reasonable to assume that any observed differences in post-test scores are probably due to the career education program.

The Career Quiz is a test of occupational knowledge developed jointly by COP's Elementary Consultant and the evaluation team's Instrumentation Specialist. It consists of 45 matching and multiple-choice items designed to measure knowledge of the working conditions of a variety of occupations and the relationship of school subjects and avocational interests to various occupations. The test booklet and scoring key are attached. Job titles appearing in the test were selected to represent all 15 U.S.O.E. occupational clusters and all levels of educational preparation. Job titles were verified in the Dictionary of Occupational Titles and a variety of materials in COP's resource center were used to verify the accuracy of the scoring key.

A portion of a COP curriculum development workshop was devoted to the review of the Career Quiz. The 12 teachers on the team were asked to evaluate each test item by answering "yes" or "no" to the following questions:

1. Does the item reflect the intent of the district's career education efforts at the elementary level?
2. Is it free of sex-stereotyping?
3. Is it free of ambiguity?
4. Are the format and reading level acceptable for at least ninety percent of fourth graders? (Only elementary teachers were asked this.)

An additional ten career educators in the state performed the same review by mail.

Any item which received more than two "no's" for any one question or more than five "no's" for all questions combined was re-written or eliminated. Fifteen original items were eliminated on the basis of question 1 and 8 were re-written on the basis of questions 2, 3, and 4. Test results are shown below as raw scores. Only students who spent the entire year in the same classroom were included in the analysis.

TABLE C

Career Quiz Results

Group	N	Std. Dev.	Mean
Career Education	60	4.39	32.22
Comparison	42	5.68	12.46

7a-e$_3$
8a$_3$

A one-tailed t-test for independent groups confirmed the significance of the rather dramatic superiority of the career education group's performance, $t(100) = 18.93$, $p < .001$. Whereas the comparison group's mean of 33% correct is only slightly above the test's "guess rate", the career education group's average score was 85% correct. Since the importance of fourth graders' possession of occupational knowledge may not be immediately obvious, the reader is encouraged to review the discussion of the COP career development model presented in the performance report.

8c$_3$

These evaluation results demonstrate that where the COP career education model was applied, it resulted in improved math achievement and career awareness and could be expected to have the same impact on other groups of similar students. Furthermore, the results do not preclude the possibility that the COP model may affect reading achievement in a setting where the reading program is in need of improvement.

8a

Even making the unlikely assumption that the only student benefits of Lincoln's career education program were those addressed in the evaluation, we find that the documented benefits were achieved at a relatively low program maintenance cost of about $5.79 per student. Included in this calculation are the costs of four field trips for 66 children at $75.00 each and $1.25 per student provided by COP to involved teachers for purchasing expendable materials.

8c

Not included in the matintenance cost are initial curriculum development, inservice training, and the purchase of durable materials. The maintenance of established programs and the introduction of the program into more schools will require continued inservice training and update of the curriculum and the materials center. This can be accomplished for the district's 12 elementary schools at an annual cost of $23,000.00 to cover both operating expenses and the salaries of a curriculum specialist and a quarter-time secretary.